

# Distilling one-qubit magic states into Toffoli states

Bryan Eastin\*

Northrop Grumman Corporation, Baltimore, MD

(Dated: February 22, 2013)

For certain quantum architectures and algorithms, most of the required resources are consumed during the distillation of one-qubit magic states for use in performing Toffoli gates. I show that the overhead for magic-state distillation can be reduced by merging distillation with the implementation of Toffoli gates. The resulting routine distills 8 one-qubit magic states directly to a Toffoli state, which can be used without further magic to perform a Toffoli gate.

Quantum algorithms frequently include a reversible classical subroutine that dominates the computation. Consequently, the Toffoli gate, which is universal for classical reversible computing, is commonly the most-used gate in an algorithm. Toffoli gates are inconvenient in many quantum architectures, but they can be implemented using, for example, one-qubit magic states and Clifford gates, where the Clifford gates are taken here to include both unitary Clifford operators and measurement and preparation in Pauli eigenbases. The initial preparation of magic states is generally poor, so prior to use, it is necessary to distill them, increasing their fidelity with the intended state. Magic-state distillation can be a significant burden. Recent quantum architecture papers indicate that when running Shor's algorithm on interesting problem sizes 90% of the physical qubits can easily be devoted to magic-state distillation [1, 2]. Such observations have helped to spur a significant body of new work focused on reducing the resources required by magic-state distillation routines [3–8].

With a few exceptions [9–11], past research on magic-state distillation has focused on routines that transform multiple faulty copies of a magic state into fewer improved copies of the same state. Being simple to inject, one-qubit magic states are a natural starting point for distillation, and routines that output the same sort of state as the input are convenient. Ultimately, however, the reason for distilling magic states is frequently to implement a Toffoli gate. Rather than segregating the two tasks, I show in this paper that one can combine them to obtain reductions in the resources required to implement a Toffoli gate.

I describe here a novel magic-state distillation routine, the  $H$ -to-Toffoli routine, that takes 8 copies of the one-qubit magic state  $|H\rangle = \cos(\frac{\pi}{8})|0\rangle + \sin(\frac{\pi}{8})|1\rangle$  that suffer  $Y$  errors with probability  $p$  and, on success, outputs a single Toffoli state,  $|T_{\text{offoli}}\rangle$ , that suffers errors with a probability of roughly  $28p^2$ . One measure of the efficiency of a distillation routine is the state cost, that is, the number of input copies of the magic state required per improved output. Using the  $H$ -to-Toffoli routine, the state cost for implementing a Toffoli gate with quadratically reduced error is competitive with the most efficient

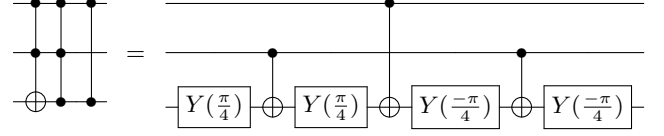


FIG. 1. Decompositions of the Margolus-Toffoli gate. (See Ref. [13].) On the left is a decomposition of the Margolus-Toffoli gate in terms of a true Toffoli gate, a controlled-controlled-sign ( $^{CC}Z$ ) gate, and a controlled-sign ( $^CZ$ ) gate. Note that the Margolus-Toffoli gate is equivalent to a Toffoli gate followed (or preceded) by the transformation  $|101\rangle \rightarrow -|101\rangle$ . On the right is a decomposition in terms of Clifford gates and  $\pi/4$  rotations about the  $Y$  axis of the Bloch sphere.

distillation routines known [7, 8]. Moreover, the location cost, which I define as the number of locations in the distillation circuit per output, is smaller by a factor of 2 to 4 for the same task.

Aside from the aforementioned differences, the scenario considered here is the standard one for magic state distillation [12]: Clifford gates are taken to be perfect while magic states are assumed to suffer from a limited (by twirling) set of errors. The notation largely follows that of Meier et al. [3].

The remainder of this paper is organized as follows: Section I introduces the  $H$ -to-Toffoli routine, and Sec. II explains how a related approach can be used for distilling Toffoli states. The efficiency of the  $H$ -to-Toffoli routine and its relative performance are discussed in Sec. III. The conclusion appears in Sec. IV, and the circuits used in the calculation of location costs are given in the appendix.

## I. $H$ -TO-TOFFOLI DISTILLATION

At the heart of the new distillation routine are two observations: First, the standard circuit for the Margolus-Toffoli gate (shown in Fig. 1), when implemented using twirled faulty  $|H\rangle$  states (see Fig. 2), is equivalent to a perfect Margolus-Toffoli gate potentially followed by  $Y$  errors on the target and  $Z$  errors on the controls, and furthermore, an error on a single  $|H\rangle$  state always results in a  $Y$  error on the target. Second, given Margolus-Toffoli gates that occasionally suffer from  $Y$  (or  $X$ ) errors on the target one can use several such gates to prepare a Toffoli

\* Bryan.Eastin@ngc.com

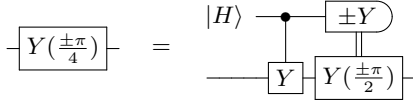


FIG. 2. Circuit identity showing how the rotation  $Y(\frac{\pm\pi}{4}) = \cos(\frac{\pi}{8})I \mp i\sin(\frac{\pi}{8})Y$  can be implemented using Clifford gates, the state  $|H\rangle$ , and the (Clifford) rotation  $Y(\frac{\pm\pi}{2})$ .

state with multiple target qubits and then check the target qubits against each other to reduce the probability of an undetected error on the prepared Toffoli state. Verified Toffoli states can then be used to implement Toffoli gates using the indirect method of Shor (shown in Fig. 3), or they can be further checked against each other.

The  $H$ -to-Toffoli magic-state distillation routine is shown in Fig. 4. Both Margolus-Toffoli gates use the same control qubits, but they have different targets. On such input the Margolus-Toffoli gate acts like the Toffoli gate (see Fig. 5), a classical reversible gate. Consequently, in the absence of errors, measuring the two target qubits in the  $Z$ -eigenbasis would yield the same result, and thus the parity measurement will yield 0. It is straightforward to show that an error on an  $|H\rangle$  state used to implement the Margolus-Toffoli gate as shown in Figs. 1 and 2 can be propagated to a  $Y$  error on the target qubit together with, possibly,  $Z$  errors on the control qubits. A single such error will thus be detected by the parity measurement while any two errors will go undetected. To lowest non-trivial order the acceptance probability  $a(p)$  is thus  $1 - 8p$  and the output error probability  $e(p)$  is  $\binom{8}{2}p^2 = 28p^2$ . Exhaustive counting yields

$$\begin{aligned} a(p) &= 1 - 8p + 56p^2 - 224p^3 + 560p^4 \\ &\quad - 896p^5 + 896p^6 - 512p^7 + 128p^8 \quad \text{and} \\ e(p)a(p) &= 28p^2 - 168p^3 + 476p^4 \\ &\quad - 784p^5 + 784p^6 - 448p^7 + 112p^8. \end{aligned}$$

Conditional on acceptance, only  $Z$  errors afflict the output control qubits, while the output target qubit is afflicted only by  $X$  errors. As it happens, each of the seven possible non-trivial errors is equally likely.

The probability of an undetected  $X$  error on the output target qubit can be made arbitrarily small by generating more target qubits and checking their parities; the  $X$ -error probability can be reduced to  $O(p^o)$  by generating  $o$  target qubits and checking them against one another. This does not reduce the probability of a  $Z$  error on the output control qubits below  $O(p^2)$ ; in fact, the coefficient of  $p^2$  worsens as  $o$  becomes larger. If further reductions in the probability of error on the control qubits are necessary, one can resort to generic Toffoli-state distillation.

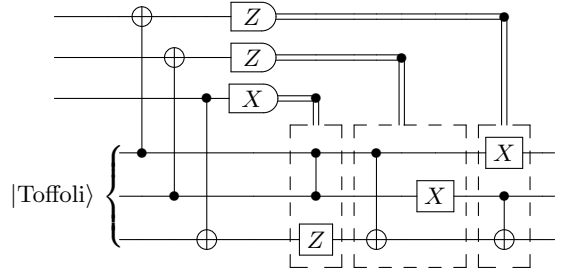


FIG. 3. A circuit implementing the Toffoli gate using a Toffoli state,  $|Toffoli\rangle = (|000\rangle + |100\rangle + |010\rangle + |111\rangle)/2$ , as per Shor [14]. The target of the Toffoli gate corresponds to the third qubit in each block.

## II. TOFFOLI-STATE DISTILLATION

For Toffoli-state distillation, it is helpful to restrict the errors that must be considered to  $Z$  errors on the control qubits and  $X$  errors on the target qubit. Conveniently, the  $H$ -to-Toffoli routine outputs states with errors of just this form. Nevertheless, should it be necessary, Aliferis has shown that the desired error model can be enforced by twirling the Toffoli state with the appropriate set of Clifford gates [15].

Toffoli states that suffer only  $X$  errors on the target qubit and  $Z$  errors on the control qubits can be used to implement Toffoli gates that suffer only  $X$  errors on the target and  $Z$  errors on the (matching) controls. Consequently, a quadratic reduction in the probability of an error on the target qubit of such a state can be achieved by implementing the left circuit of Fig. 4 using Toffoli states. This distillation routine can be shown to be equivalent to the Toffoli-state distillation routine proposed by Aliferis [15]. Given identically prepared Toffoli states, the probability of an  $X$  error on the target is reduced from  $p$  to roughly  $p^2$  and the probability of errors that do not involve an  $X$  error on the target roughly doubles. Reduction of the  $Z$ -error probability on a control qubit can be achieved using the same circuit if one first swaps the target qubit with the control qubit of the Toffoli state using a pair of Hadamard gates (see Fig. 5). This transformation takes  $Z$  errors on the former control qubit to  $X$  errors on the new target qubit and vice versa.

## III. EFFICIENCY

The  $H$ -to-Toffoli routine is atypical of magic-state distillation routines in that its inputs and outputs are different and, as a consequence, it is not composable with itself. This complicates comparisons with other distillation routines. Taking the Toffoli state to have a “value” of 4  $|H\rangle$  states, it might be said that the  $H$ -to-Toffoli routine costs 2 input  $|H\rangle$  states per output  $|H\rangle$  state with quadratically reduced error probability, numbers which correspond to a scaling exponent of  $\log_2 2 = 1$ . This

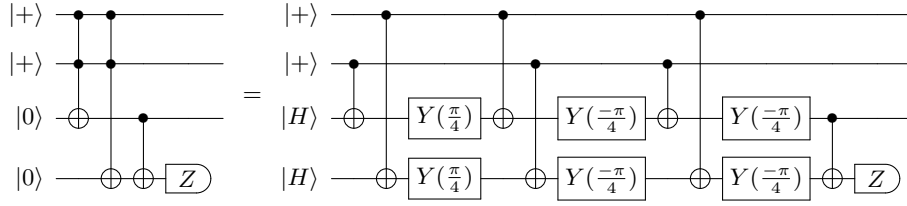


FIG. 4.  $H$ -to-Toffoli distillation circuit. The output is discarded whenever a non-trivial measurement outcome is obtained. The circuit on the left shows the distillation in terms of Toffoli gates while the circuit on the right shows the same distillation circuit expanded in terms of the  $|H\rangle$ -state implementation of Margolus-Toffoli gates. All  $Y(\pm\frac{\pi}{4})$  gates are implemented indirectly as in Fig. 2. By enumeration and error propagation it is easily shown that, to lowest non-trivial order, this circuit takes  $|H\rangle$  states that suffer  $Y$  errors with probability  $p$  to Toffoli states that suffer errors (some combination of  $X$  errors on the target qubit and  $Z$  errors on the controls) with probability  $28p^2$ .

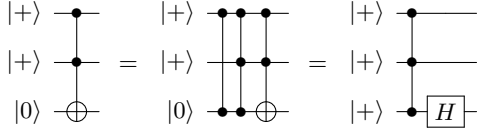


FIG. 5. Toffoli state preparation. The circuit on the left shows an obvious method of preparing the Toffoli state. The middle circuit shows Toffoli-state preparation using the Margolus-Toffoli gate (decomposed into three more familiar gates). From the right circuit it is clear that the target qubit of a Toffoli state can be changed using a pair of Hadamard gates. The first equality follows from the fact that a gate controlled on  $|0\rangle$  is not executed; the same logic implies that a Margolus-Toffoli can be substituted for a true Toffoli gate whenever the target qubit is initially prepared in the state  $|0\rangle$ , as will be the case whenever the Margolus-Toffoli gate is used in the distillation routines presented here.

would make it competitive with the most efficient distillation routines [7, 8], but the comparison is unfair in two ways: The output is not really a collection of  $|H\rangle$  states, and the routine is not scalable, being fixed in size and non-composable. For this reason, I consider below the overhead required for specific tasks: the production of a Toffoli state or Toffoli gate using faulty  $|H\rangle$  states, where the ( $|H\rangle$ -state) inputs and (state or gate) outputs suffer errors with probability  $p$  and  $O(p^2)$ , respectively.

Using the  $H$ -to-Toffoli routine, 8  $|H\rangle$ -type magic states which suffer  $Y$  errors with probability  $p$  are required to distill a single Toffoli state which suffers errors with probability  $O(p^2)$ . As illustrated in Fig. 3, Clifford gates and a Toffoli state suffice to perform a Toffoli gate, so a Toffoli gate can be implemented with the same parameters. I assume in this analysis that all Toffoli gates are performed using Toffoli states. For distillation routines other than the  $H$ -to-Toffoli routine,  $|H\rangle$  states are distilled prior to being used in the Toffoli-state preparation circuit shown in Fig. 6. Using this circuit, only 4  $|H\rangle$  states are required to prepare each Toffoli state and therefore to implement each Toffoli gate. Bravyi and Haah have shown that  $|H\rangle$  states with quadratically suppressed errors can be prepared at a cost arbitrarily close to 3 input  $|H\rangle$  states

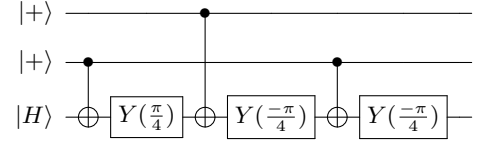


FIG. 6. Toffoli-state preparation circuit based on the Margolus-Toffoli gate. This circuit can be used, together with circuits from Figs. 3 and 2, to implement a Toffoli gate using only 4  $|H\rangle$  states, as opposed to 7  $|H\rangle$  states as commonly assumed in the literature.

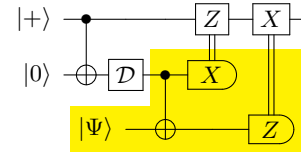


FIG. 7. (Color online) State injection. The circuit shown injects an arbitrary state  $|\Psi\rangle$  into a quantum code. The unshaded portion of the circuit is implemented on encoded qubits, while the shaded gates are performed on unencoded qubits. The gate  $\mathcal{D}$  represents decoding the quantum code.

per output  $|H\rangle$  state [7]. Jones has further shown that quadratic error suppression can be obtained at a state cost arbitrarily close to 2 as part of a larger distillation routine [8]. Multiplying each of these numbers by 4, one finds that the  $H$ -to-Toffoli routine yields an improvement of 33% in the state cost compared to the best routines of Bravyi and Haah and performs similarly to Jones' routines.

In addition to the state cost, I calculate location costs for the  $H$ -to-Toffoli distillation routine and some close competitors. Locations are simply points in (discretized) space and time where a qubit is undergoing a gate or storing quantum information. The location cost of a distillation routine is the number of locations required per output. In an effort to make the cost less dependent on the native gate set, one-qubit unitary Clifford gates are ignored when counting locations; one-qubit Clifford measurements are also ignored on the grounds that these

Distillation routine	State cost	Output error probability	Location cost	
			Toffoli state	Toffoli gate
10-to-2	20	$36p^2$	183	298
14-to-2	28	$28p^2$	179	334
26-to-6	17.33	$76p^2$	151	252.7
$H$ -to-Toffoli	8	$28p^2$	36	91

TABLE I. Properties of Toffoli-gate and Toffoli-state implementations based on a selection of distillation routines. The 14-to-2 and 26-to-6 routines are members of a family of routines developed by Bravyi and Haah [7]. Within that family the 26-to-6 routine seems to have the lowest location cost. The 10-to-2 routine is a relatively efficient routine proposed in Ref. [3]. The circuits on which the location counts for these distillation routines are based appear in the appendix. The location costs quoted for the Toffoli gate include both the circuit in Fig. 3 and the cost of state injection.

tend to be much faster than two-qubit Clifford gates on robustly encoded qubits.

As illustrated by Fig. 8 in the appendix, the number of locations in the  $H$ -to-Toffoli distillation routine is 36. The number of additional locations required to implement the Toffoli gate as shown in Fig. 3 using the resultant Toffoli state is 15. The total resource cost for implementing a Toffoli gate is thus 51 locations and 8  $|H\rangle$  states. If each  $|H\rangle$  state is injected from a lower level of encoding using the circuit in Fig. 7 then 5 additional locations are required per  $|H\rangle$  state and the total location cost rises to 91 locations per Toffoli gate. By comparison, the appendix contains a compacted circuit for the 26-to-6 routine that requires only 53.7 locations per improved  $|H\rangle$  state counting state injection and 32 discounting it. The circuit in Fig. 6 requires 23 locations and 4  $|H\rangle$  states. Thus, using the 26-to-6 routine, 151 locations are required to produce a Toffoli state with quadratically reduced error and 252.7, including state injection, for a similarly improved Toffoli gate. The  $H$ -to-Toffoli routine reduces the location cost for Toffoli states and gates by a factor of 4.2 and 2.8, respectively. State injection is included in the latter number as an indication of the true cost of a Toffoli gate. This represents the lower end of the possible location overhead per input  $|H\rangle$  state;  $|H\rangle$  states having previously undergone distillation will cost many more locations, thereby enhancing the relative performance of the  $H$ -to-Toffoli routine.

The properties of Toffoli-state and Toffoli-gate implementations using some of the most resource-efficient distillation routines are summarized in Tab. I.

#### IV. CONCLUSION

Magic-state distillation, as required for the implementation of robust Toffoli gates, is a significant driver of overhead in some proposed quantum computing architectures. In this paper I have shown that the overhead can be reduced by merging magic-state distillation with the implementation of Toffoli gates. I described a novel distillation routine that distills 8 copies of a one-qubit magic state,  $|H\rangle$ , directly into a Toffoli state with quadratically reduced error probability. For the purpose of implement-

ing Toffoli gates, the state cost of this routine is as good as or better than existing routines and the location cost is a factor of 2 to 4 lower compared to other distillation routines that provide quadratic error suppression.

Subsequent to the development of the  $H$ -to-Toffoli routine described herein, I learned that Cody Jones has developed a closely related routine, albeit one with a distinct motivation and a very different looking quantum circuit. Jones' routine likewise distills 8 copies of a one-qubit magic state into a single Toffoli state and provides the same amount of error suppression. More information on this complementary work can be found in Ref. [16].

My analysis has focused on a single round of magic-state distillation. Partly this reflects the fact that the proposed routine outputs Toffoli states, for which few distillation routines are known, and partly it reflects a belief that many rounds of distillation at some fixed level of encoding will be uncommon, as has previously been argued in Refs. [7, 17]. Distillation routines that provide a quadratic reduction in error per round are among the most efficient [7], and it is unnecessary to perform distillation using encoded gates that are significantly less error prone than the outputs of the distillation routine [5, 7, 17], so it seems likely that maximum efficiency will typically be achieved by doing magic-state distillation at many different levels of encoding, where a quadratic reduction in the error probability is achieved at each level. As illustrated in the recent architectural paper of Fowler et al. [2], surface codes are well suited to this approach, being very flexible in the degree of encoding. For concatenated coding, the steps in the degree of encoding will typically be greater, so routines with stronger error suppression may play a role, but given the difference in the sizes of the coefficients for the output error probabilities for magic-state distillation as opposed to fault-tolerant error correction, I would not expect very high-order distillation to be necessary.

I have introduced the notion of location cost because I feel that this more accurately identifies the resource that one wishes to minimize in order to control the overhead for fault-tolerant quantum computing. In the limit of many rounds of distillation, the state cost becomes the determining factor in the location cost, but when only a few rounds of distillation are required the two can

yield significantly different rankings of distillation routines. The location cost is not a particularly portable measure; the best method of counting locations will vary between different architectures, as is illustrated by a recent paper optimizing the location cost of the 15-to-1 magic-state distillation routine for a surface code architecture [18]. Additionally, while I have focused on the implementation of the Toffoli gate, it is perhaps worth mentioning that the Margolus-Toffoli gate can be implemented at a slightly lower location cost for some of the distillation routines considered here using the circuit in Fig 1. As has been previously noted, the Margolus-Toffoli gate is often an acceptable stand-in for the Toffoli gate [19]; in particular, the Margolus-Toffoli gate can be used in a quantum computation to calculate any classical function that is subsequently exactly undone, an occurrence not uncommon in reversible computing.

On the topic of magic-state distillation, several interesting open problems remain. Thus far, Toffoli-state distillation has received little attention. The only published routine has a state cost of 8 for quadratic error suppression [15], though a more efficient method was recently suggested by Jones in the conclusion of Ref. [1]. An obvious avenue of investigation is thus to search for new

routines distilling Toffoli states. The compaction of existing distillation routines to minimize their location costs in various architectural settings is another worthwhile endeavor. Aside from magic-state distillation, other (older) techniques exist for implementing non-Clifford gates in concatenated codes [14] and there even exist quantum codes, including some varieties of surface code, that require no distillation at all [20, 21]. One might therefore investigate the performance tradeoff between magic-state distillation and these other techniques for implementing fault-tolerant non-Clifford gates. Finally, despite the caveats listed above, it would be interesting to determine the location cost as a function of the target error rate for a variety of distillation sequences under the assumption of perfect Clifford gates.

## ACKNOWLEDGMENTS

I am grateful to Adam Meier, Manny Knill, Ben Reichardt, Kevin Obenland, and Andrew Landahl for enlightening discussions on this topic. Special thanks go to Cody Jones for pointing me to other routines for the distillation of Toffoli states. The magic-state distillation routines described herein are patent pending.

- 
- [1] N. Cody Jones, Rodney Van Meter, Austin G. Fowler, Peter L. McMahon, Jungsang Kim, Thaddeus D. Ladd, and Yoshihisa Yamamoto, “Layered architecture for quantum computing,” *Phys. Rev. X* **2**, 031007 (2012), [arXiv:1010.5022](#).
  - [2] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland, “Surface codes: Towards practical large-scale quantum computation,” *Phys. Rev. A* **86**, 032324 (2012), [arXiv:1208.0928](#).
  - [3] Adam M. Meier, Bryan Eastin, and Emanuel Knill, “Magic-state distillation with the four-qubit code,” *Quant. Inf. Comp.* **13**, 0195–0209 (2013), [arXiv:1204.4221](#).
  - [4] Jonas T. Anderson, “On the power of reusable magic states,” (2012), [arXiv:1205.0289](#).
  - [5] Tomas Jochym-O’Connor, Yafei Yu, Bassam Helou, and Raymond Laflamme, “The robustness of magic state distillation against errors in Clifford gates,” *Quant. Inf. Comp.* **13**, 0361–0378 (2013), [arXiv:1205.6715](#).
  - [6] Earl T. Campbell, Hussain Anwar, and Dan E. Browne, “Magic state distillation in all prime dimensions using quantum Reed-Muller codes,” *Phys. Rev. X* **2**, 041021 (2012), [arXiv:1205.3104](#).
  - [7] Sergey Bravyi and Jeongwan Haah, “Magic state distillation with low overhead,” *Phys. Rev. A* **86**, 052329 (2012), [arXiv:1209.2426](#).
  - [8] Cody Jones, “Multilevel distillation of magic states for quantum computing,” (2012), [arXiv:1210.3388](#).
  - [9] Ben W. Reichardt, “Quantum universality from magic states distillation applied to CSS codes,” *Quant. Inf. Proc.* **4**, 251–264 (2005), [arXiv:quant-ph/0411036](#).
  - [10] Ben W. Reichardt, “Quantum universality by state distillation,” *Quant. Inf. Comp.* **9**, 1030–1052 (2009), [arXiv:quant-ph/0608085](#).
  - [11] Earl T. Campbell, “Catalysis and activation of magic states in fault-tolerant architectures,” *Phys. Rev. A* **83**, 032317 (2011), [arXiv:1010.0104](#).
  - [12] Sergey Bravyi and Alexei Kitaev, “Universal quantum computation with ideal Clifford gates and noisy ancillas,” *Phys. Rev. A* **71**, 022316 (2005), [arXiv:quant-ph/0403025](#).
  - [13] Guang Song and Andreas Klappenecker, “Optimal realizations of simplified Toffoli gates,” *Quant. Inf. Comp.* **4**, 361–372 (2004), [arXiv:quant-ph/0312225](#).
  - [14] Peter W. Shor, “Fault-tolerant quantum computation,” in *Proc. 37th Symp. Found. Comp. Sci.* (1996) p. 56, [arXiv:quant-ph/9605011](#).
  - [15] Panos Aliferis, “Level reduction and the quantum threshold theorem,” (2007), [arXiv:quant-ph/0703230](#).
  - [16] Cody Jones, “Novel constructions for the fault-tolerant Toffoli gate,” (2012), [arXiv:1212.5069](#).
  - [17] Robert Raussendorf, Jim Harrington, and Kovid Goyal, “Topological fault-tolerance in cluster state quantum computation,” *New Journal of Physics* **9**, 199 (2007), [arXiv:quant-ph/0703143](#).
  - [18] Austin G. Fowler and Simon J. Devitt, “A bridge to lower overhead quantum computation,” (2012), [arXiv:1209.0510](#).
  - [19] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter, “Elementary gates for quantum computation,” *Phys. Rev. A* **52**, 3457–3467 (1995), [arXiv:quant-ph/9503016](#).



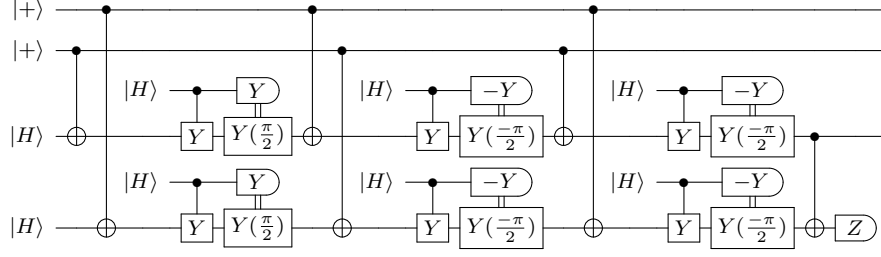


FIG. 8.  $H$ -to-Toffoli distillation circuit, expressed using exclusively Clifford gates and  $|H\rangle$  states. In total, this circuit uses 8  $|H\rangle$  states and 36 locations to distill a single Toffoli-state output. For the purpose of counting locations, measurements and unitary one-qubit Clifford gates are ignored.

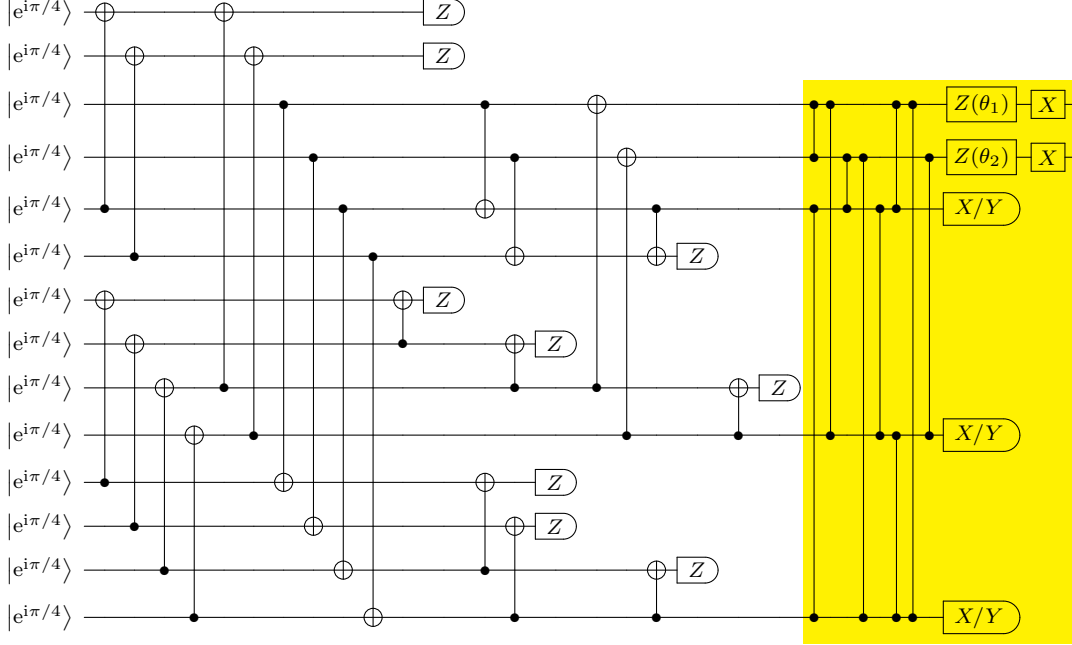


FIG. 9. (Color online) Compacted circuit for the 14-to-2 distillation routine [7] expressed in terms of Clifford gates and  $|e^{i\pi/4}\rangle$  states. The  $C_Z$  gates in the highlighted region act on every pair of unmeasured qubits. Depending on the  $Z$ -measurement outcomes, some subset of the highlighted gates are implemented. The angles  $\theta_1$  and  $\theta_2$  are multiples of  $\pi/2$ , where the multiple is likewise dependent on the  $Z$ -measurement outcomes. The output is discarded whenever the outcome of any of the remaining measurements is non-trivial. Using standard circuit identities, it can be shown that any distillation routine of the sort proposed by Bravyi and Haah in Ref. [7] can be expressed in this form, though efficiency-wise it is not always desirable. In total, this circuit uses 14  $|e^{i\pi/4}\rangle$  states and at most 78 locations to distill 2 improved copies of  $|e^{i\pi/4}\rangle$ . Conditional on success, the marginal probability of error is reduced from  $p$  to roughly  $7p^2$ . For the purpose of counting locations, measurements and unitary one-qubit Clifford gates are ignored. Note that  $|e^{i\pi/4}\rangle = (|0\rangle + e^{i\pi/4}|1\rangle)/\sqrt{2} = e^{-i\pi/8} H Z(\frac{-\pi}{2})|H\rangle$  so the states  $|e^{i\pi/4}\rangle$  and  $|H\rangle$  are equivalent for the purpose of location counting.

- [20] Emanuel Knill, Raymond Laflamme, and Wojciech Zurek, “Accuracy threshold for quantum computation,” (1996), [arXiv:quant-ph/9610011](https://arxiv.org/abs/quant-ph/9610011).
- [21] Hector Bombin and Miguel A. Martin-Delgado, “Topological computation without braiding,” *Phys. Rev. Lett.* **98**, 160502 (2007), [arXiv:quant-ph/0610024](https://arxiv.org/abs/quant-ph/0610024).

## Appendix: Location counting circuits

The location costs quoted for distillation routines in this paper were obtained using the circuits shown in Figs. 8, 9, 10, and 11.

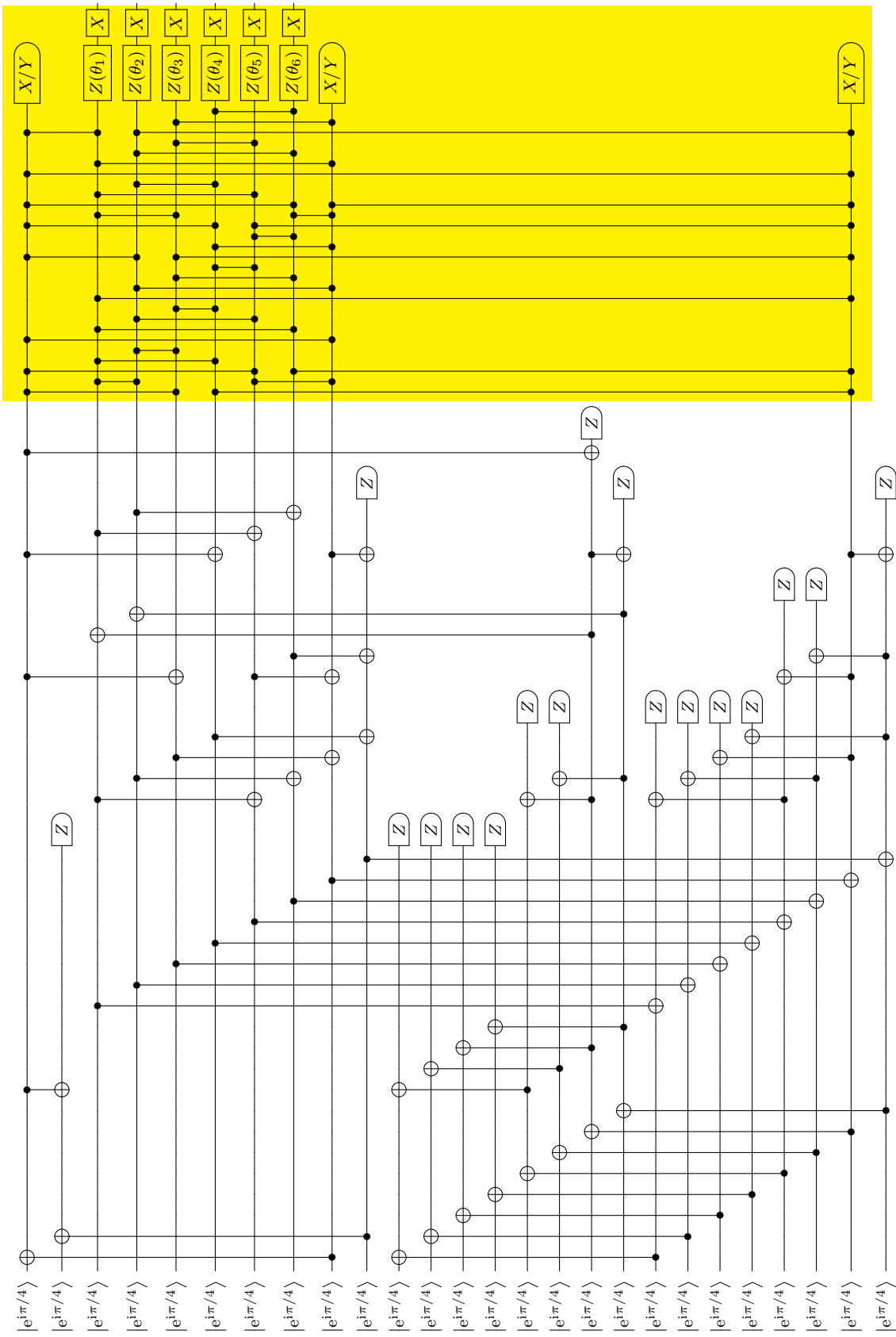


FIG. 10. (Color online) Compacted circuit for the 26-to-6 distillation routine [7] expressed in terms of Clifford gates and  $|e^{i\pi/4}\rangle = (|0\rangle + e^{i\pi/4}|1\rangle)/\sqrt{2}$  states. Depending on the Z-measurement outcomes, some subset of the highlighted gates are implemented. The angles  $\theta_i$  are multiples of  $\pi/2$ , where the multiple is likewise dependent on the Z-measurement outcomes. The output is discarded whenever the outcome of any of the remaining measurements is non-trivial. In total, this circuit uses 26  $|e^{i\pi/4}\rangle$  states and at most 192 locations to distill 6 improved copies of  $|e^{i\pi/4}\rangle$ . Conditional on success, the marginal probability of error is reduced from  $p$  to roughly  $76p^2$ . For the purpose of counting locations, measurements and unitary one-qubit Clifford gates are ignored.

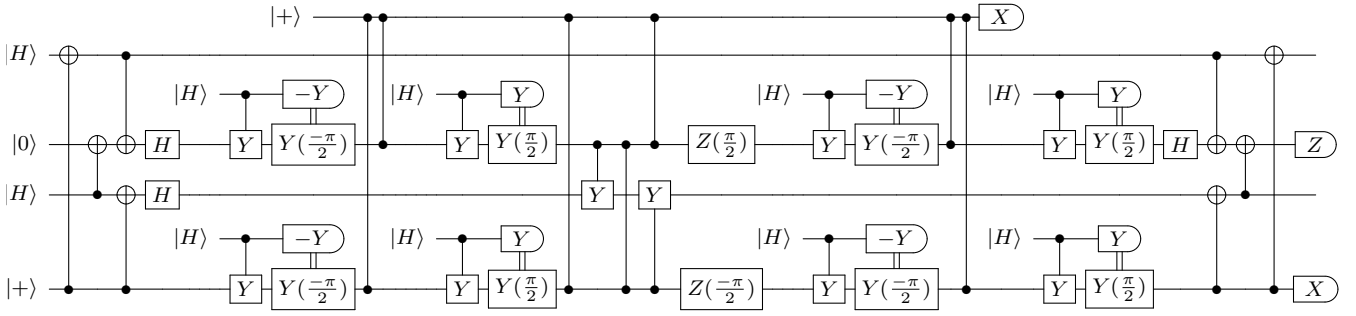


FIG. 11. Circuit for the 10-to-2 magic-state distillation routine [3] expressed in terms of Clifford gates and  $|H\rangle$  states. The output is discarded whenever a non-trivial measurement outcome is obtained. In total, this circuit uses 10  $|H\rangle$  states and 80 locations to distill 2 improved copies of  $|H\rangle$ . Conditional on success, the marginal probability of error is reduced from  $p$  to roughly  $9p^2$ . For the purpose of counting locations, measurements and unitary one-qubit Clifford gates are ignored.